Midterm Project

107061517 張簡崇堯

Outline

- The MobileNet
- Hardware Design
- SCML Platform Results

MobileNet

- Convolution operations: 1254 times.
- In layer 3, the input feature size is 17 * 17.
- Convolution with size > 17 * 17 : 223 times.

Hardware Design

- Decompose 3D depthwise convolution to 2D convolution.
- The input size is set to 17 since the input size of most convolution operations are smaller than 17.

Design Specification

- I/O port, clock, reset
- Three 2D local buffers (assume stride = 2)
  1. Input feature: 17 * 17, 32-bit unsigned int
  2. Input partial sum: 8 * 8, 32-bit unsigned int
  3. Convolution kernel: 3 * 3, 32-bit unsigned int

1st Synthesis Result

- Use sim_V_BASIC
  - Runtime: 27300 ns
  - Area: 201945.4
- Use sim_V_DPA
  - Runtime: 21540 ns
  - Area: 196520.3

After Loop Unrolling

- Use sim_V_BASIC
  - Runtime: 8340 ns
  - Area: 195660.9
- Use sim_V_DPA
  - Runtime: 4170 ns
  - Area: 207568.7

After Loop Unrolling & Flatten Array

- Use sim_V_BASIC
  - Runtime: 3550ns
  - Area: 279850
- Not use sim_V_DPA. Since HLS takes time!

Synthesis Issue

- When synthesizing stride = 1 with loop unrolling, data hazards will happen.
- To solve this issue, the loop unrolling is limited to kernel only.

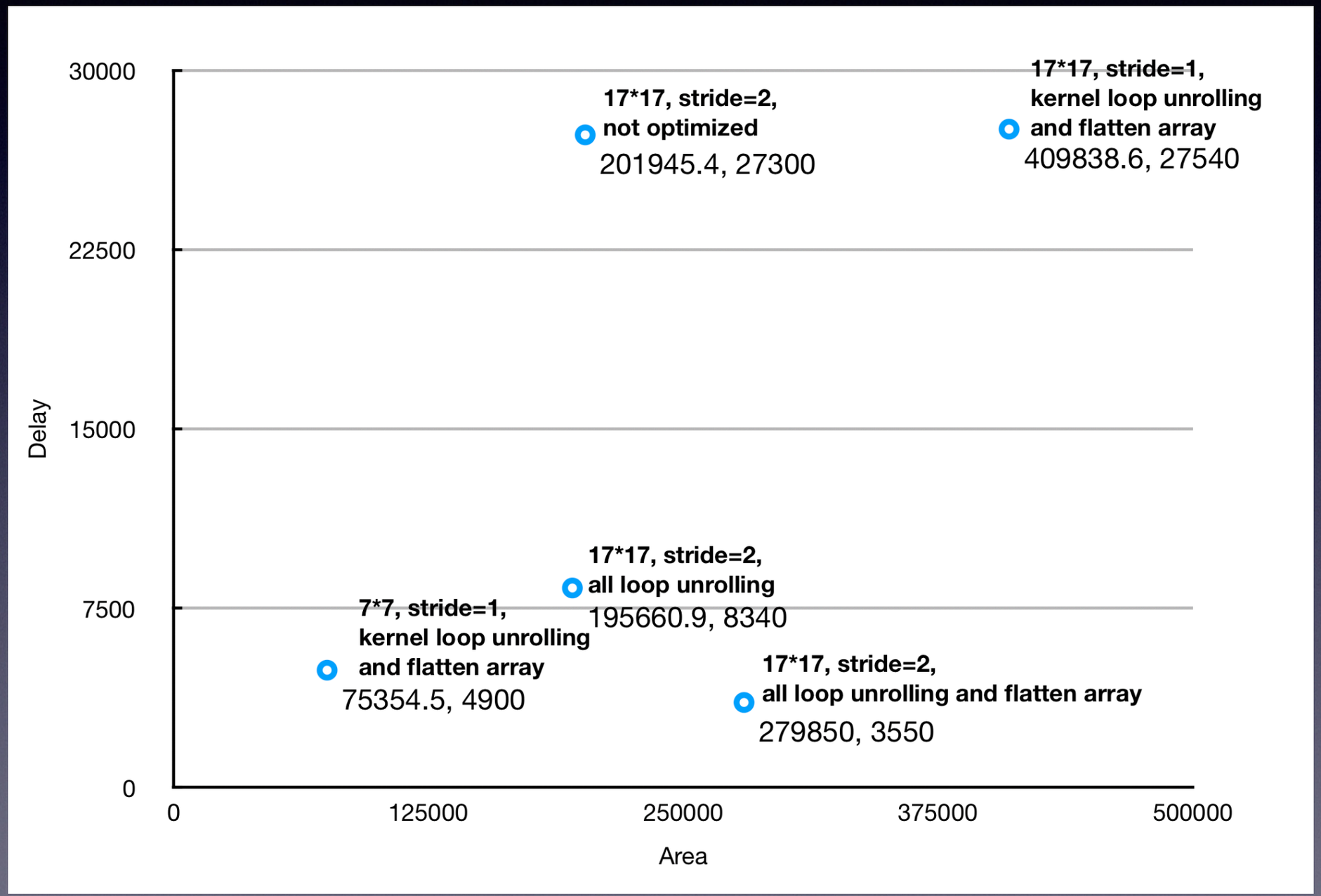
Convolution with stride = 1

- Use sim_V_BASIC
- Runtime: 27540 ns
- Area: 409838.6

Optimization for convolution with size = 7 * 7

- I noticed that there are many convolution operations with size = 7 in the MobileNet.
- Maybe we can build another 7 * 7 engine with stride = 1 to deal with them.
- Use sim_V_BASIC
  - Runtime: 4900 ns
  - Area: 75354.5

Design Comparison



SCML Platform Results

- Use 17*17 engine only:
  5435489115 ns

- Use 17*17 and 7*7 (for stride=1) engine:
  3703649115 ns